Business and Economic Research
ISSN 2162-4860
2016, Vol. 6, No. 2

Determinants of Baseball Success: An Econometric Approach

Jacob Andrew Loree

Ryerson University，Canada

E-mail: jacob.loree@ryerson.ca

Received: May 20, 2016 Accepted: June 7, 2016

doi:10.5296/ber.v6i2.9488 URL: http://dx.doi.org/10.5296/ber.v6i2.9488

Abstract

While much has been investigated into the relationship between several baseball statistics and success, the literature is more heavily focused on individual level characteristics and the salary of individual baseball players. This paper investigates, at a more macro level, the importance of key baseball statistics on the level of wins a team can expect on average using the Lahman Baseball Database for all teams from 1985 to 2015. After several robustness tests, the most important variables an average team should focus on is the total number of runs a team gives up and getting on base as often as possible (by walks as well as base hits). The paper finds that while team salary is statistically significant, it takes an unreasonably large change in salary to be meaningful in terms of number of wins recorded. Therefore, previous research on the effect of salary on team success may be overblown.

Keywords: Baseball Statistics, Econometrics, Sports

1. Introduction

Baseball has been a fertile ground for statistical investigation in the past thirty years. This is due to the large amount of data available, as well as the general consistency of rules and team make-up year over year. Unlike other sports, teams are rarely able to go from “first to last” in a division in a single season in baseball. Instead, they must build up a strong team. Compared to other major sports, baseball is the most competitive in terms of equality of teams (Ben-Naim et al, 2006). This means that the level of outliers in baseball data is relatively low and the data is relatively stationary. Through the time period investigated (1985-2014), the number of games played yearly was 162[1] where 81 are played at home and the other 81 played in other team’s cities. While injuries are quite common in baseball, a stretch of long-term injuries to several players on one team is unlikely. Therefore there are less random

[1] Other than 1994, where labour relations broke down and caused a strike and shortened season.

shocks that affect gameplay. Unlike other sports, losing a very good player for a section of the season is less impactful – due to the longer schedule and more team-based gameplay.

As well, baseball generates an enormous amount of money in America. In 2014, Major League Baseball (MLB) generated $9 billion dollars in gross revenue (Brown, 2014) while Americans spent over $615 million on equipment to play themselves (Active Marketing, 2007). With such money involved, determining how exactly teams can win the most games, and by extension draw more fans and money to them is of paramount importance.

As described by Michael Lewis in the book *Moneyball,*

"The sheer quantity of brain power that hurled itself voluntarily and quixotically into the search for new baseball knowledge was either exhilarating or depressing, depending on how you felt about baseball. The same intellectual resources might have cured the common cold, or put a man on Pluto." (Lewis, 2003)

The concept of being able to discern team success through statistics was first put forward by Bill James in his Baseball Abstract (James, 1987). While decades have passed and more complex and formal statistical techniques have been used, research has yet to reach a consensus on what helps teams succeed. Recently in the literature (somewhat due to the importance of *Moneyball*), the concept of team salary has been argued as a main driver of team success. Both anecdotally and empirically, this argument has been a factor in baseball policy in the creation of a luxury tax for high spending teams, which is redistributed to teams with low payrolls. This paper will attempt to clarify which baseball statistics have a strong and important relationship with team wins. By adding new data at the team level in several different baseball statistics (including team salary) we hope to help clarify the relationship between several statistics and their effect on team success to help add focus to where management should strengthen their teams.

2. Relevant Literature

Overall, the field of research relating baseball statistics to victories in baseball is relatively small. The consensus of this field is that strong pitching and good fielding are essential to building a strong team, while offense is less important in most metrics (Fullerton, Fullerton and Walke, 2014). However, these studies usually focus on a restricted sample of only a few or even one season. When the sample is extended further, the significance of these metrics diminishes. In particular, when team salary is casually tested to influence performance, there is very little impact across a whole sample of 1980-2000, however there is strong significance when tested only in the 1990's (Hall, Syzmanski and Zimbalist, 2002). Therefore, it is possible for the significance of variables to impact team success differently in different time periods, which needs to be taken into account when specifying the model.

Competitive balance is an important issue in any sport and as such, baseball has been intensely analyzed to determine the competitiveness of all teams. Competitive balance refers to how often the same teams are in the very top of the league. A less competitively balanced league would have dynastic teams – teams that win a lot, win more each consecutive year. This is seen almost universally as bad. Fans of baseball want surprise and novelty and one way for that to happen is new teams winning. This competitive balance can be thought of in economic theory terms. In general, we want additional competition between firms as this leads to increases in social welfare. The level of competitive balance in baseball is of some argument in the literature. Some studies have found that the level of competition has increased, leading to closer games and more attendance (Schmidt and Berri, 2001). Interestingly, Schmidt and Berri (2004) also found that this increase in competitive balance has lead to "clusters of convergence". That is, while overall competitive balance has increased, there are specific tiers of teams that have clustered together. This means that while the difference in wins may have decreased overall, there are distinct levels of team success that teams rarely move out of. This paper will test competitive balance by regressing a team's level of wins on the wins it achieved last year. If competitive balance is strong, we should expect the autoregressive coefficient to be insignificant or very small.

Most studies employ a White heteroskedatic methodology in order to correct the variance within the error term (Gustafson and Hadley, 2007), but some find that the variance of the error is white noise (Keener, 2013). Since there is conflict in the literature, regressions using robust standard errors will be run alongside regressions excluding these standard errors. If there are significant differences between the two, additional specification tests should be run.

Most of the literature in previous years have focused on the **salary hypothesis**. That is, teams that are able to spend more, on average, see more wins per season. This hypothesis was first put forward by the Commissioner's Blue Ribbon Panel on Baseball Economics, which found that there were large disparities between the payrolls of teams, and that these discrepancies impact competitive balance (Levin, et al. 2000). Most studies in this field find at least partial support for the hypothesis holding (Hasan, 2008).

Related to the salary hypothesis, many researchers have investigated the free agency market and how it affects teams. Since the salary hypothesis is such an important portion of this literature, determining how to spend that money on players is also paramount. There are two distinct hypothesizes that relate team success and wage dispersion. That is, there are two separate theories on spending all of your money on a single high level player or many cheaper, low level players and how it will effect success. Firstly, there is the "danger potential" hypothesis (Ramasway and Rowthorn, 1991). The danger potential hypothesis as it relates to baseball is that wage disparity is naturally occurring due to different labour types. These types are inherent to players and players cannot imitate types. These types relate to how well a player coexists on a team. Players who do nor "play nice" must be paid more to not actively harm team productivity. As such, there is natural wage dispersion not related to team ability. Secondly, there is the team cohesion hypothesis presented by David Levine (1991). This states that wage disparity leads to lower team performance, since there is envy and infighting amongst the team due to differences in wages. A study by Craig Depken (1999, 2000) found strong evidence of the Levine team cohesiveness hypothesis and in general the literature finds that a more divisive payroll leads to lower wins otherwise (Kahn, 1993).

While we do not have player level data in terms of salaries, this idea of wage dispersion is important to keep in mind when discussing policy implications for improving teams. For example, if homeruns are found to be significant to team success, signing one powerful player may have a different effect than signing many cheap players – even if the change in homeruns is identical.

3. Methodology

This paper will look at several baseball statistics to determine which are statistically and economically significant to a baseball team's success. This will be accomplished using a few different methodologies – all which use the same variables.

We investigate both a fixed effect model[2] and a standard OLS model. The model specification is:

$$Wins_{it} = \alpha + \pi Wins_{it-1} + \beta batting_{it} + \gamma pitching_{it} + \delta fielding_{it} + \theta salary_{it} + \varphi AL + \rho_t + \tau_i + \varepsilon_{it} \quad (1)$$

Appendix 1 explains all variables. The error term either assumes homoscedasticity or allows for heteroskedasticity based on estimation strategy. This model estimates the effect of a matrix of batting, pitching, and fielding statistics on a team's success. As well, we test the salary hypothesis. We include time and team fixed effects to help control for any unexplained fluctuations year to year as well as unexplained team attributes. This is important in a sport like baseball, as it is not uncommon for a team (or player) to play far above or below expectations for a period of time. Controlling for these variations in a long sample help see the true effect of these statistics.

We also estimate a log-log model to determine the effect of percentage changes in statistics on the percentage change on wins, which we call the **elasticity of success**. This model is estimated identically, except all variables are transformed into their log versions.

As well, since there seems to be evidence of significance changing based on time periods, we estimate the full sample and also partial samples. These partial samples are simply either half of the sample period, using the same estimation strategy. Here, we are seeing if 1980-2000 has a significantly different relationship between performance and wins than 2000-2014.

4. Data

All data comes from the Lahman Baseball Database, a database provided freely to econometricians. The entire database includes all available statistics on teams going back into the late 1800's, before the formation of Major League Baseball. This database has been culled to include all teams in Major League Baseball, and is a sample period of 1985-2014. This range was chosen as this was the earliest data available for team salary – a vital statistic for our estimation. Table 1 below demonstrates the summary statistics for the data.

[2] A fixed effect model is chosen due to a rejection of a random effect test using a Hausmann test

Business and Economic Research
ISSN 2162-4860
2016, Vol. 6, No. 2

Table 1. Summary Statistics

Variable	Mean	Standard Deviation	Min	Max
Wins	79.68	11.94	43	116
Hits	1429.95	109.19	963	1684
Doubles	275.14	34.75	159	376
Triples	30.93	8.96	11	61
Home Runs	157.03	37.02	58	264
Walks	527.26	74.99	319	775
Strikeouts	1034.99	152.91	568	1535
Stolen Bases	105.11	36.61	25	314
Runs Against	731.55	96.15	448	1103
Complete Games	9.58	7.48	0	47
Errors	109.55	19.96	54	179
Double Plays	147.62	19.79	82	204
Fielding Percentage	0.98	0.005	0.97	0.99
Team Salary	5750000	4100000	88000	232000000
% of Teams in American League	0.49	0.49	0	1

As one can see, there are not a large number of outliers in the statistics, other than team salary. When comparing team salary, additional influences like inflation and media revenue can push player salaries upwards. This can be partially controlled with time fixed effects, but we may not see as clear a picture of the relationship between salary and performance as possible.

While hits are made up of some of the other variables involved in the estimation (namely doubles, triples, and home runs), these “extra base” hits are inherently more valuable than just an additional hit. Therefore, it is worth separating out the different kinds of hits to see if there are different effects. As well, different hitters are more likely to hit different types of hits[3]. As seen in figure 1, these extra base hits only make up a small fraction of total hits.

[3] This is the argument between “contact” hitters and “power” hitters. Power hitters are more likely to hit extra base hits – but usually cost much more than contact hitters.

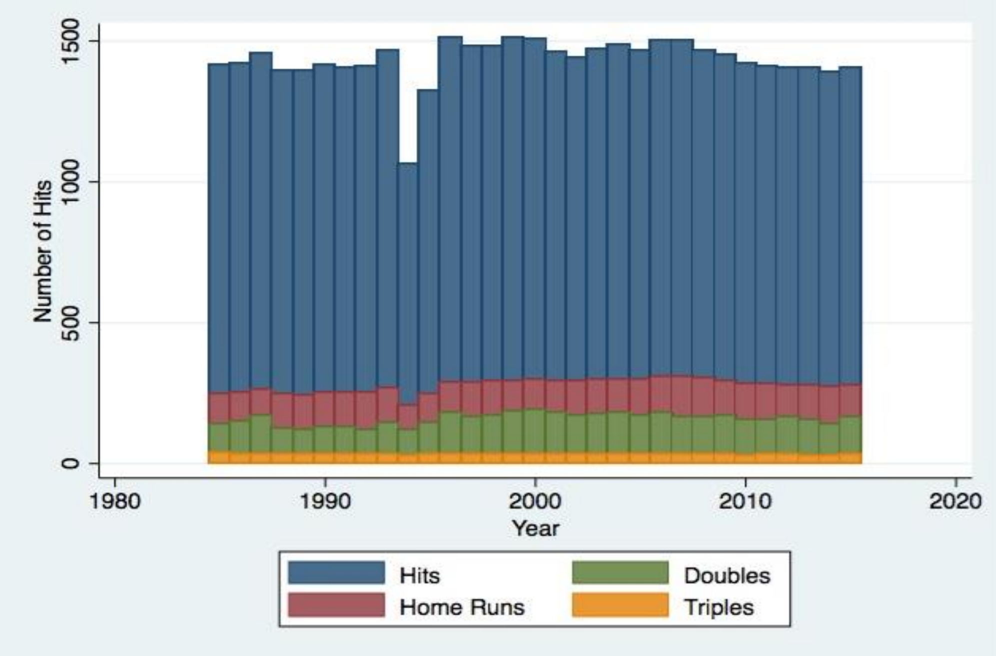

Figure 1. Average Hits per Year by Type of Hit

Table 2 gives the hypothesized relationships between wins and the different statistics. Most of these are intuitive to the sport of baseball. For example, hitting more homeruns will most likely lead to more victories. Less intuitively, we hypothesize that a lagged version of wins from last season will have a positive effect on wins this season. This is because, while baseball is considered very competitive in relation to other sports, a good team (holding everything else equal) will continue being successful into the next season. While this is not always the case due to retirements and players moving teams, we believe there is positive autocorrelation of victories for a team.

Table 2. Hypothesized Relationships to Wins

Variable	Hypothesized Effect on Wins
Lagged Wins	Positive
Hits	Positive
Doubles	Positive
Triples	Positive
Homeruns	Positive
Walks	Positive
Strikeouts	Negative
Stolen Bases	Positive
Runs Against	Negative
Complete Games	Positive
Errors	Negative
Double Plays	Positive
Fielding Percentage	Positive
AL	Positive

5. Results[4]

The first result we find is that regardless of specification, the results are robust. Regardless of fixed effects, robust standard errors, or variables specified, we receive similar significant point estimates. Table 3 and 4 give regression results for level changes and percentage changes respectively. The interpretation for Table 3 is a one unit change in the explanatory variable will lead to the coefficient change in victories. The interpretation for Table 4 is a one percent change in the explanatory variable will lead to a coefficient percentage change in victories. Both regressions are done to more easily demonstrate the intuition. Table 3 measures team salaries in thousands of dollars, since one dollar changes lead to extremely small changes in wins. In this same vein, fielding percentage is measured as a one percent change (0.01 as opposed to 1) as it is measured between 0 and 1.

There are several interesting things that can be seen from these regressions. Firstly, regardless of specification, most variables have the relationship with success that we hypothesized. As well, most of the statistics we investigate hold at least some statistical significance. However, many of these variables' point estimates are so small that when combined with average changes in those variables they are, in effect, minute changes in the number of wins. This is an important point, as these regressions seem to say that all statistics hold at least some importance to determining wins, but depending on how much a statistic evolves, it may actually be a nonfactor. After averaging across years and teams, the average change in variables are essentially zero. That is, there is very little trend year over year.

Table 3. Unit Change Regressions

	(1)	(2)	(3)	(4)	(5)
Lagwin	0.32	0.004	-0.022	-0.03	-0.03
	(0.005)***	(0.002)	(0.003)***	(0.004)***	(0.004)***
Hits	0.04	0.06	0.048	0.045	0.045
	(0.0008)***	(0.0004)***	(0.0007)***	(0.001)***	(0.001)***
Double	-0.02	0.002	0.01	0.012	0.012
	(0.002)***	(0.001)*	(0.001)***	(0.002)***	(0.002)***
Triple	0.083	0.032	0.034	0.088	0.08
	(0.007)***	(0.004)***	(0.004)***	(0.006)***	(0.006)***
Homerun	0.05	0.09	0.11	0.12	0.12
	(0.001)***	(0.001)***	(0.001)***	(0.002)***	(0.002)***
Walk		0.039	0.036	0.043	0.04
		(0.0004)***	(0.0005)***	(0.001)***	(0.0007)***
Strikeout		0.003	-0.002	-0.004	-0.004
		(0.0002)***	(0.0004)***	(0.0006)**	(0.0006)***
Stolen Base		0.029	0.023	0.037	0.037
		(0.0008)***	(0.0009)***	(0.001)***	(0.001)***

[4] All results show robust standard errors. The differences between significance and point estimates when estimations are run with or without robust standard errors are extremely small.

Run Against		-0.101	-0.103	-0.099	-0.1
		(0.0004)***	(0.0004)***	(0.001)***	(0.001)***
Complete Game		0.05	0.012	0.027	0.029
		(0.005)***	(0.006)**	(0.007)***	(0.006)**
Error		0.061	0.01	0.004	0.005
		(0.002)***	(0.003)***	(0.004)	(0.005)
Double Play		0.002	-0.018	-0.012	-0.01
		(0.002)	(0.001)***	(0.002)***	(0.002)***
Fielding Percentage		2.75	1.31	1.61	1.61
		(1.08)***	(0.12)***	(0.13)***	(0.13)***
Team Salary (in thousands)	0.00002	0.000002	0.000007	-0.00002	0.000002
	(0.000001)***	(0.000001)**	(0.000001)***	(0.0000005)***	(0.000005)***
AL		0.6	-5.8	9.96	9.65
		(0.06)***	(0.39)***	(0.89)***	(0.31)***
Sample	Full	Full	Full	1985-2000	2000-2014
Year and Team Fixed Effects	No	No	Yes	Yes	Yes
Adjusted R^2	0.38	0.86	0.86	0.88	0.85

Standard errors are in parenthesis. ***,**,* indicate 1,5,10% significance respectively.

Each equation includes a constant.

Table 4. Percentage Change Regressions

	(1)	(2)	(3)	(4)	(5)
Lagwin	0.95	0.72	0.68	0.72	0.63
	(0.01)***	(0.002)***	(0.003)***	(0.018)***	(0.019)***
Hits	0.53	0.35	0.27	0.22	0.29
	(0.004)***	(0.005)***	(0.07)***	(0.008)***	(0.018)***
Double	-0.004	0.013	0.03	0.027	0.05
	(0.002)	(0.003)***	(0.003)***	(0.004)***	(0.006)***
Triple	0.002	0.004	0.005	0.01	0.006
	(0.0008)***	(0.0007)***	(0.0008)***	(0.001)***	(0.0014)***
Homerun	0.009	0.05	0.072	0.063	0.091
	(0.001)***	(0.001)***	(0.003)***	(0.004)***	(0.005)***
Walk		0.08	0.08	0.09	0.081
		(0.001)***	(0.002)***	(0.006)***	(0.006)***
Strikeout		0.007	-0.015	-0.017	-0.011
		(0.001)***	(0.003)***	(0.004)**	(0.005)**
Stolen Base		0.01	0.01	0.013	0.006
		(0.0006)***	(0.0007)***	(0.001)***	(0.001)***
Run Against		-0.29	-0.32	-0.28	-0.39

		(0.003)***	(0.004)***	(0.018)***	(0.02)***
Complete Game		0.0007	-0.002	-0.001	0
		(0.0.0003)**	(0.0003)***	(0.0006)	(0.0006)
Error		0.022	0.006	0.009	-0.01
		(0.002)***	(0.002)**	(0.003)***	(0.01)
Double Play		-0.0006	-0.012	0.0001	-0.03
		(0.001)	(0.002)***	(0.002)	(0.003)***
Fielding Percentage		0.915	0.612	0.657	-0.152
		(0.077)***	(0.09)***	(0.09)***	(0.734)
Team Salary	0.001	-0.001	0.0004	-0.0012	0.005
	(0.0003)***	(0.0004)***	(0.0007)	(0.0012)	(0.001)***
AL		0.002	-0.03	0.001	-0.03
		(0.0004)***	(0.004)***	(0.002)***	(0.006)***
Sample	Full	Full	Full	1985-2000	2000-2014
Year and Team Fixed Effects	No	No	Yes	Yes	Yes
Adjusted R^2	0.94	0.95	0.95	0.96	0.96

Standard errors are in parenthesis. ***,**,* indicate 1,5,10% significance respectively.

Each equation includes a constant.

Instead, teams must go out and sign and attract better players to realize these gains. Splitting the sample into two periods seems to not demonstrate two extremely significant relationships. However, in regards to fielding percentage, there is a shift in the relationship. Pre-2000, increases in fielding percentage led to more victories. Post-2000, it decreased the number of wins a team could expect. We hypothesize that this means that there is a nonlinear relationship between victories and fielding percentage. After 2000, fielding percentages are so high that the only way to push them higher is sacrificing another aspect of the game (batting, speed, etc.). This is an avenue for future research.

Surprisingly, there seems to be very little significance, statistical or otherwise, in regards to wins in a given season and wins in the previous season. This was a hypothesis we expected to very easily be proven by our estimations. Instead, the evidence is scattered and even at its least well defined (column 1 of table 3 and 4) very weak. While there is a significant positive relationship between wins in one season and the next, the relationship is relatively muted. This seems to suggest that there is a strong competitive balance to Major League Baseball. That is, holding all else equal, a team with many wins will continue having many wins, but slightly less than before. We find little evidence of **winning momentum** – a concept that explains that a team will continue winning due to the fact it was winning previously. Therefore, a strong team cannot rest on its laurels, as without improving, the level of success they achieve will slowly backslide. As such, there is fairly strong evidence of competitive balance.

The other biggest surprise of our regressions is the lack of strong evidence for the salary hypothesis. Significance and sign of the relationship seems to shift based on estimation

Business and Economic Research
ISSN 2162-4860
2016, Vol. 6, No. 2

strategy. As well even at its strongest, it would take a very large increase in team salary to induce a meaningful change in team victories. An average team's payroll increased by $3175 over the entire sample period. That means that a team would have to spend 16000 times more than a usual change to win one additional game purely through the income effect of victories[5]. While the effect of increased hitting, pitching, and fielding is not contained within the salary coefficient, we can say that increasing a team's payroll without regard to how the team will perform is not an appropriate way to success. Therefore, we reject the strictest form of the salary hypothesis – that a team can find success purely through additional spending. However, since our estimations find that better players (as defined as more hits or less strikeouts, etc) contribute to victories and these better players cost more, a weaker version of the salary hypothesis still may be valid.

6. Conclusion

In conclusion, we do not find evidence of either winning momentum, nor of the salary hypothesis in its strictest form. We do find statistical significance for a large number of our variables, and in general they fit the hypothesized relationship. A team can become more successful if they can hit more, limit the number of runs a team can score against them, and successfully field hit balls in the field. While this is basic intuition that a team most likely already knows, our estimation helps prioritize which of these statistics has the largest impact on wins. By our estimations, additional homeruns and decreasing runs the opposing team scores are the best avenues to increase wins. However, strikeouts are not nearly as harmful as originally hypothesized. A team would have to strikeout an additional 500 times over a 162 game season to lose 1 additional game – holding everything else equal[6]. While not as important as the homerun, walks are more significant than originally thought. From table 2, we see that walks happen almost twice as often as doubles and nearly 18 times as often as triples. However, in terms of creating victories, walks are much more significant than doubles and almost as significant as triples. Therefore, having a team draw more walks seems like a relatively easy way to generate success.

This paper has investigated several different statistics in order to find the determinants of baseball victories. Our estimation finds that strong pitching and fielding lead to more wins. While there is additional room for research – specifically in the nonlinearity of fielding percentage on wins – the estimations seem to be robust. However, using these estimation strategies, we find no evidence of the strongest version of the salary hypothesis or strong evidence of winning momentum. In the future, this research can be improved by investigating predictions of individual player's statistics. If teams are better able to determine future performance of players, they can use the above estimations more accurately. We hope that this paper has helped to clear up issues in the literature.

References

Active Marketing Group. (2007). The Baseball Market.

---

[5] Calculation done by author.
[6] From column 3 of table 3.

Ben-Naim, E., Vazquez, F., & Redner, S. (2006). Parity and predictability of competitions. *Journal of Quantitative Analysis in Sports*, 2(4). http://dx.doi.org/10.2202/1559-0410.1034

Brown, Maury. (2014, December 10th). Major League Baseball Sees Record $9 Billion in Revenues for 2014. *Forbes*.

Depkin, C. (1999). Free-Agency and Competitiveness of Major League Baseball. *Review of Industrial Organization*, *14*(3). http://dx.doi.org/10.1023/A:1007788303098

Depkin, C. (2000). Wage Disparity and Team Productivity: Evidence from Major League Baseball. *Economics Letters, 67*(1). http://dx.doi.org/10.1016/S0165-1765(99)00249-9.

Fullerton, S., Fullerton, T, & Walke, A. (2014). An Econometric Analysis of the 2013 Major League Baseball Season. *Research in Business and Economics Journal, 9*.

Gustafson, E., & Hadley, L. (2007). Revenue, Population, and Competitive Balance in Major League Baseball. *Contemporary Economic Policy, 25*(2). http://dx.doi.org/10.1111/j.1465-7287.2007.00036.x

Hall, S., Szymanski, S., & Zimbalist, A. (2002). Testing Casuality Between Team Performance and Payroll: The Cases of Major League Baseball and English Soccer. *Journal of Sports Economics*, *3*(2), 149-168. http://dx.doi.org/10.1177/152700250200300204.

Hasan, S. (2008). Can Money Buy Success?: A Study of Team Payroll and Performance in the MLB. *Journal of Global Business Management, 4*(2).

James, B. (1987). *The Bill James Baseball Abstract 1987*. Ballantine Books.

Kahn, L. (1993). Free-Agency, Long Term Contracts, and Compensation in Major League Baseball. *The Review of Economics and Statistics, 75*(1). http://dx.doi.org/10.2307/2109641.

Keener, M. (2013). The econometrics of baseball: A statistical investigation. *Research in Business & Economics Journal, 9*.

Sean Lahman. (2015). *Lahman Baseball Database.* Retrieved from http://www.seanlahman.com/baseball-archive/statistics/

Levin, R. C., Mitchell, G. J., Volcker, P. A., & Will, G. F. (2000). The report of the independent members of the commissioner's blue ribbon panel on baseball economics, Major League Baseball.

Levine, D. (1991). Cohesiveness, Productivity and Wage Dispersion. *Journal of Economic Behaviour and Organization, 15*. http://dx.doi.org/10.1016/0167-2681(91)90031-R.

Ramaswamy, R., Rowthorn, R. (1991). Efficiency Wages and Wage Dispersion. *Economica, 58*. http://dx.doi.org/10.2307/2554695

Schmidt, M. & Berri, D. (2001). Competitive Balance and Attendance: The Case for Major League Baseball. *Journal of Sports Economics, 2*(2). http://dx.doi.org/10.1177/152700250100200204.

Schmidt, M. & Berri, D. (2004). Convergence and Clustering in Major League Baseball: The Haves and Have Nots? *Applied Economics, 36*. http://dx.doi.org/10.1080/0003684042000261851.

Appendix 1. Description of Variables

Variable	Description
Wins	Number of wins a team accomplishes in a season
Batting	A matrix of batting statistics that number of total hits, doubles, triples, homeruns, walks, stolen bases, and strikeouts.
Pitching	A matrix of pitching statistics that includes number of runs against and complete games pitched by one pitcher
Fielding	A matrix of fielding statistics that includes percentage of plays fielded successfully and number of double plays performed and errors
Salary	Measure of total team spending on all active players on a roster
AL	Dummy variable indicating if a team plays in the American League, as opposed to the National League

Glossary

Complete Game – A game pitched entirely by a single pitcher.

Double Play – A fielding play in which two outs are recorded on a single at bat.

Error – Recorded when a fielder misplays a relatively simple ball. Awarded by the official scorer’s judgement.

Number of Runs Against – Total number of runs of all opposing teams across an entire season.

Strikeouts – In this paper, strikeouts are defined as number of at bats a batter on a team strikes out (achieving three strikes). It is not a count of the number of strikeouts induced by a team’s pitcher. Therefore, it has a negative correlation with scoring runs.